TECHNICAL REPORT

Estimation and Optimal Control for Constrained Markov Chains

by D.-J. Ma, A.M. Makowski and A. Shwartz

ISR TR 86-40

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **DEC 1986**	2. REPORT TYPE	3. DATES COVERED **00-00-1986 to 00-00-1986**

4. TITLE AND SUBTITLE
Estimation and Optimal Control for Constrained Markov Chains

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Maryland,Electrical Engineering Department,College Park,MD,20742

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**24**	

Invited paper to the 25th IEEE Conference on Decision and Control,
Athens, Greece, December 1986

ESTIMATION AND OPTIMAL CONTROL FOR CONSTRAINED MARKOV CHAINS

by

D.-J. Ma[1], A. M. Makowski[1] and A. Shwartz[2]

University of Maryland and Technion

ABSTRACT

The (optimal) design of many engineering systems can be adequately recast as a Markov decision process, where requirements on system performance are captured in the form of constraints. In this paper, various optimality results for constrained Markov decision processes are briefly reviewed; the corresponding implementation issues are discussed and shown to lead to several problems of parameter estimation. Simple situations where such constrained problems naturally arise, are presented in the context of queueing systems, in order to illustrate various points of the theory. In each case, the structure of the optimal policy is exhibited.

I. INTRODUCTION

Controlled queueing systems constitute a natural class of models for a variety of engineering applications, such as the ones arising in computer communication networks and in manufacturing environments [9,10,23]. The theory of Markov decision processes (MDP's) provides one of the main tools to analyze performance optimality for many such stochastic systems. In recent years, these methods were successfully used on several simple queueing systems, to obtain an explicit form for the optimal control strategy [8,12,25,31]. The model considered here

[1] Electrical Engineering Department and Systems Research Center, University of Maryland, College Park, Maryland 20742, U.S.A. The work of these authors was supported partially through NSF Grant ECS-83-51836, partially through a grant from AT&T Bell Laboratories and partially ONR Grant N00014-84-K-0614.

[2] Department of Electrical Engineering, Technion, Haifa 32000, Israel. The work of this author was supported through a Grant from AT&T Bell Laboratories.

is one operating in *discrete time*; this reflects the increasing use of digital implementation in many of today's systems; moreover, under standard assumptions the analysis of some classes of continuous-time MDP's reduces to the analysis of an embedded discrete-time MDP [16,27]. A large body of knowledge is available in the literature on this class of problems, ranging from optimality conditions to structural properties of the optimal strategies to algorithmic results. The discussion is typically carried out under the assumption that system performance may be captured through a *single* cost criterion based on an instantaneous cost which depends on the state, the control action and perhaps time. Standard performance measures include the *finite-horizon* criterion, as well as the *discounted* and *long-run average* costs, both being taken over the infinite horizon.

However, in many practical situations, conflicting goals need to be taken into consideration, a requirement that often *cannot* be adequately captured through a single cost function. Typical examples arise in computer communication applications where it may be desirable to maximize throughput, while keeping delays small, or in manufacturing systems where resources are allocated so as to maximize the so-called line throughput and yet prevent inventories to build-up.

Such trade-off considerations require that the standard formulation for MDP's be modified so as to accomodate the conflicting objectives. One possible way to achieve this would be to define several cost functions, one for each objective identified by the designer, and to focus attention on the corresponding multi-objective optimization problem. Here instead, conflicting requirements inherent to the (optimal) design problem at hand are incorporated through constraints. Such an approach was considered by Ross [26] in the context of finite-state MDP's under an objective criterion and a constraint function, each being given in the form of a long-run average functional asociated with an instantaneous cost, and the goal is to maximize the first criterion subject to a bound on the constraint. As will become apparent from the discussion given in forthcoming sections, constrained MDP's form a rich class of stochastic optimization problems whose solution leads to a variety of interesting questions of both a theoretical and practical nature.

This paper, which is devoted to a brief survey of recent work in this area, is organized as follows: In Section 2, the single constraint optimization problem is precisely formulated, a solution technique via Lagrangian arguments is then outlined and general results on the structure of optimal policies are summarized. Extensions to more complex (i. e., non-Markovian) dynamics as well as to the more difficult situation where several constraints are enforced, are briefly considered in Section 3. Also discussed in Section 3 are various implementation issues which are inherent to the very form of the optimal policies identified in Section 2, and which lead naturally to specific problems of combined estimation and control for Markov chains. Most of the literature on this topic [11] is concerned with *indirect* adaptive control problems [7]. For constrained MDP's, the situation is somewhat different in that some control parameters may not be available to the decision-maker in practice, even if the model parameters were known. This suggests viewing the design of implementable constrained optimal policies as a *direct* adaptive control problem [7]. The various ideas and proposals of Section 3 are illustrated in two specific situations of independent interest, which are discussed in Sections 4 and 5, respectively. The first situation centers around a problem of resource sharing, where several discrete-time queues with geometric service requirements compete for the service attention of a *single* server. The second example is essentially a problem of optimal flow control for a discrete-time $M|M|1$ queue.

2. MDP'S WITH A SINGLE CONSTRAINT

2.1 The problem formulation

In [26], Ross considers the following version of the constrained optimization problem for Markov chains: Let $\{X(n)\}_1^\infty$ denote a controlled Markov chain, with *countable* state space S, *compact* metric action space U and transition probabilities $(p_{xy}(u))$ assumed *continuous* in u. Following the usual formulation of MDP's as given in [13], an *admissible* policy π generates at time n an action $U(n)$ on the basis of the information $H(n) := (X(1), U(1), \cdots, X(n-1), U(n-1), X(n))$. For a given initial state distribution (held fixed hereafter), the policy π induces a probability measure P^π on the natural σ-field that equips the canonical sample space $\Omega := (S \times U)^\infty$, with corresponding expectation operator E^π. The notation $\mathcal{P}$ is reserved for the collection of all admissible policies. The class of (possibly randomized) Markov stationary policies is then denoted by $\mathcal{F}$, while $\mathcal{G}$ stands for the subclass of all non-randomized policies in $\mathcal{F}$. Clearly $\mathcal{G} \subset \mathcal{F} \subset \mathcal{P}$.

Given are two mappings $r, c: S \times U \to \mathbb{R}$, which are assumed *continuous* in the variable u and which are interpreted as the instantaneous *reward* and *cost* functions, respectively. For every admissible policy π in $\mathcal{P}$, pose

$$J(\pi) := \underline{\lim}_n \frac{1}{n} E^{\pi} \sum_{t=1}^{n} r(X(t), U(t)) \tag{2.1}$$

and

$$K(\pi) := \overline{\lim}_n \frac{1}{n} E^{\pi} \sum_{t=1}^{n} c(X(t), U(t)), \tag{2.2}$$

and for every V in $\mathbb{R}$, define

$$\mathcal{P}_V := \{\pi \text{ in } \mathcal{P} : K(\pi) \leq V\}. \tag{2.3}$$

Of interest here is the constrained problem $(C\mathcal{P}_V)$ defined as

$$(C\mathcal{P}_V): \quad \text{maximize } J(\pi) \text{ over } \mathcal{P}_V.$$

2.2 A Lagrangian methodology

A Lagrangian methodology is now described for studying this constrained problem; it requires the introduction of a family of auxiliary problems: For every $\gamma > 0$, let the mapping $b^{\gamma}: S \times U \to \mathbb{R}$ be given by

$$b^{\gamma}(x, u) := r(x, u) - \gamma c(x, u) \tag{2.4}$$

for all (x, u) in $S \times U$, and define the corresponding Lagrangian functional by

$$B^{\gamma}(\pi) := \underline{\lim}_n \frac{1}{n} E^{\pi} \sum_{t=1}^{n} b^{\gamma}(X(t), U(t)) \tag{2.5}$$

for every policy π in $\mathcal{P}$. As it occurs in Mathematical Programming, the solution of the constrained MDP $(C\mathcal{P}_V)$ is closely related to various properties of the *unconstrained* MDP (LP_{γ}) associated with (2.5), where

$$(LP_{\gamma}): \quad \text{maximize } B^{\gamma}(\pi) \text{ over } \mathcal{P}.$$

To see this, observe that for *any* policy π in $\mathcal{P}$, the inequality $B^{\gamma}(\pi) \geq J(\pi) - \gamma K(\pi)$ holds, whereas if the policy π yields (2.1) and (2.2) as *limits*, then $B^{\gamma}(\pi) = J(\pi) - \gamma K(\pi)$. If, in addition to this property, a policy π^* *meets* the constraint, i. e., $K(\pi^*) = V$, and is *optimal* for the MDP (LP_γ), then necessarily for *all* π in $\mathcal{P}$,

$$B^\gamma(\pi^*) = J(\pi^*) - \gamma K(\pi^*) \geq B^\gamma(\pi) \geq J(\pi) - \gamma K(\pi). \tag{2.6}$$

Since $K(\pi) \leq V$ for any policy π in $\mathcal{P}_V$, the inequality (2.6) and the fact $\gamma > 0$ readily imply that

$$J(\pi^*) \geq J(\pi) + \gamma[K(\pi^*) - K(\pi)] = J(\pi) + \gamma(V - K(\pi)) \geq J(\pi) \tag{2.7}$$

for every policy π in $\mathcal{P}_V$, whence the policy π^* solves the constrained optimization problem $(C\mathcal{P}_V)$.

From this discussion, it should be clear to the reader in what sense the Lagrangian problems $\{(LP_\gamma), \gamma > 0\}$ are useful for solving the original constrained problem. Indeed, as the arguments given above indicate, *any* policy π^* in $\mathcal{P}$ which

(R1): yields the expressions $J(\pi^*)$ and $K(\pi^*)$ as *limits*,

(R2): meets the constraint with $K(\pi^*) = V$, and

(R3): solves the *unconstrained* MDP (LP_γ) for some $\gamma > 0$,

necessarily solves the constrained problem $(C\mathcal{P}_V)$. This approach can be used either directly on specific problems *mutatis mutandis*, as illustrated in Sections 4 and 5, or it can provide a convenient theoretical framework for establishing general results on the existence and structural form of solutions to the constrained optimization problem.

To simplify the discussion, it is convenient to assume that S is *finite* and that the controlled chain has a *single ergodic* class under each policy f in $\mathcal{F}$. In that case, under any policy f in $\mathcal{F}$, the expressions (2.1), (2.2) and (2.5) exist as *limits* and are *independent* of the initial condition. Moreover, it is well known that an optimal policy for problem (LP_γ) can always be selected to be a pure strategy in $\mathcal{G}$. This follows by standard arguments based on the corresponding Dynamic Programming equation (2.8), which states here that for all x in S, the relation

$$B^\gamma + h^\gamma(x) = \max_{u \epsilon U}[b^\gamma(x, u) + \sum_{y \epsilon S} p_{xy}(u) h^\gamma(y)] \tag{2.8}$$

holds for some real constant B^γ and some mapping $h^\gamma : S \to \mathbb{R}$. Their existence is guaranteed under the assumed conditions [13], with B^γ identified with the optimal value of problem (LP_γ).

It is also well known that if $\mathcal{G}^\gamma$ denotes the class of policies g^γ in $\mathcal{G}$ with the property that for each x in S, the action $u = g^\gamma(x)$ attains the maximum in the Dynamic Programming equation (2.8), then any policy in $\mathcal{G}^\gamma$ is optimal for the problem (LP_γ).

2.3 Optimality via randomized policies

In [26], it is shown under the simplifying assumptions stated earlier, that whenever the problem is "feasible", there exists a constrained optimal policy with a simple structure.

Theorem [26]: *If for some $0 < \gamma < \infty$, at least one policy g^γ in $\mathcal{G}^\gamma$ has the property that $K(g^\gamma) \leq V$, then there exists a constrained optimal policy f^* in $\mathcal{F}$ defined by a simple randomization between two pure Markov stationary policies $\underline{g}$ and $\overline{g}$ in $\mathcal{G}$.*

More precisely, there exist $0 < \gamma^* < \infty$, and policies $\underline{g}$ and $\overline{g}$ in $\mathcal{G}^{\gamma^*}$ such that

$$K(\underline{g}) \leq V \leq K(\overline{g}). \tag{2.9}$$

If the *randomized* policies $f_q, 0 \leq q \leq 1$, are defined by

$$f_q := q\overline{g} + (1-q)\underline{g}, \tag{2.10}$$

then $f^* \equiv f_{q^*}$, where the optimal bias q^* is determined as the solution to the equation

$$K(f_q) = V, \quad 0 \leq q \leq 1. \tag{2.11}$$

The discussion of this result can be summarized as the search for a policy in $\mathcal{F}$ that satisfies the requirements (R1)-(R3) for some value $\gamma^* > 0$ of the Lagrangian parameter. The reader is referred to [26] for details.

2.4 Optimality via mixing policies

The existence of policies $\underline{g}$ and $\overline{g}$ in $\mathcal{G}^{\gamma^*}$ with the property (2.9) can be further exploited to generate an alternate solution to the constrained MDP (CP_V) in the one-parameter family of *mixing* policies $\{\pi(p), 0 \leq p \leq 1\}$. For every p in the unit interval $[0,1]$, consider a two-sided coin biased so that the events Head and Tail occur with probability p and $1-p$, respectively. To define the mixing policy $\pi(p)$, throw this biased coin exactly *once* at the beginning of times, before starting to operate the system. The policy $\pi(p)$ is defined as the policy in $\mathcal{P}$ that operates according to $\underline{g}$ (resp. $\overline{g}$) if the outcome is Tail (resp. Head). It is not difficult to see that for such a policy, the relations

$$J(\pi(p)) = (1-p)J(\underline{g}) + pJ(\overline{g}) \tag{2.12a}$$

and

$$K(\pi(p)) = (1-p)K(\underline{g}) + pK(\overline{g}) \tag{2.12b}$$

hold true.

Now, under the non-degeneracy assumption $K(\underline{g}) \neq K(\overline{g})$, pose $p^* := [V - K(\underline{g})]/[K(\overline{g}) - K(\underline{g})] \geq 0$, (owing to (2.9)). Observe from (2.12) that the policy $\pi(p^*)$ steers (2.2) to the value V, yields (2.1) and (2.2) as *limits* and that $B^{\gamma^*}(\pi(p)) = (1-p)B^{\gamma^*}(\underline{g}) + pB^{\gamma^*}(\overline{g}) = B^{\gamma^*}$. In other words, the policy $\pi(p^*)$ meets the requirements (R1)-(R3), and consequently solves the constrained problem (CP_V). It should be noted, in contrast with the policy f^* obtained by randomization in Section 2.3, that the evaluation of the optimal mixing parameter p^* is immediate, if the values $K(\underline{g})$ and $K(\overline{g})$ are available.

3. MORE ON CONSTRAINED OPTIMIZATION

3.1 Generalizations

It is possible to generalize the discussion of Section 2 in several directions:

Firstly, it is of interest to consider system dynamics where more complex probabilistic mechanisms are allowed for state transitions and/or where the state processes live in more general spaces. Beutler and Ross [4] consider a version of (CP_V) for general semi-Markov decision processes with finite state space and compact action space. Nain and Ross [21] study a specific constrained MDP with countable (non-finite) state space. In both cases, similar results on the structure of the constrained optimal policy are reported. However, more work seems needed as no general theory is available to date in the case of *non-finite* state spaces.

Secondly, as pointed out in the introduction, the main practical motivation for studying constrained optimization problems arises from the desire to handle situations with multiple (conflicting) objectives. The next natural step would consist in formulating constrained MDP's with *multiple* constraints as a *nonlinear programming* problem in the space of policies. More precisely, with the notation of Section 2, let $J(\pi)$ be the cost associated with the policy π in $\mathcal{P}$, and let $K^i(\pi)$ denote the corresponding value of the i^{th} constraint, $1 \leq i \leq I$. For every vector $V = (V_1, \cdots, V_I)$ in $\mathbb{R}^I$, pose

$$\mathcal{P}_V := \{\pi \text{ in } \mathcal{P} : K^i(\pi) \leq V_i, 1 \leq i \leq I\} \tag{3.1}$$

and define the multiple constraint problem (CP_V) as in Section 2. Very little is known on the existence and structure of optimal solutions for problems of this type. This probably could be traced back to the fact that the relationship of the corresponding Lagrangian problem to the original problem (CP_V) is far more subtle in the multiple constraint situation. In fact, as of the writing of this paper, it is not clear on how to obtain in general an optimal policy through randomization and/or mixing procedures similar to the ones presented in Section 2. Results are available only in particular instances; Altman and Shwartz [1] establish existence of a solution to a problem with multiple constraints for the competing queue model considered by Nain and Ross [21].

3.2 Implementation issues

Even if the policies $\underline{g}$ and $\overline{g}$, and the value γ^* were readily available, the computation of the optimal bias q^* may prove to be a non-trivial task, for it requires solving for q in the *implicit* equation $K(f_q) = V$ on the interval $[0,1]$, and makes it necessary to evaluate the expression $K(f_q)$ for each $0 \leq q \leq 1$. Both steps usually turn out to be highly difficult ones in many applications, and are often possible only via numerical methods; this difficulty is clearly illustrated on the competing queue model discussed in Section 4.

The optimal bias q^* acts as an *internal* parameter; it is available *in principle* if the *external* (or model) parameters (i. e., the entries in the transition matrix) are known, but may not be easily available *in practice* owing to computational difficulties. Of course, this difficulty of implementation is further compounded when some of the external parameters are not known, since "on line" identification of the external parameters does not provide a feasible means to evaluate q^*. In any case, this points to adaptive methods for *directly* estimating q^*, now treated as an unknown parameter, and this specifically for the purpose of generating an optimal control; in the terminology of adaptive systems, this is referred to as *direct* adaptive control [7]. This suggests broadening the notion of adaptive control for Markov chains to view it as a procedure for *recursively updating the control to meet the performance criterion.* Although this is a well-known problem in the general theory of adaptive systems, it seems to have not been studied much in the context of MDP's, at least to the authors' knowledge.

The reader's attention should be drawn to the fact that direct adaptive control ideas, with q^* regarded as an unknown parameter, do not always lead to implementable policies. This was illustrated by Shwartz and Makowski [28,30] on the competing queue problem of Section 4.

The implementation issues discussed above can be addressed in the somewhat more general context of steering the cost (2.2) to a prespecified value V: Given is a parametrized family $\{f_q, 0 \leq q \leq 1\}$ of Markov stationary policies, and assume, with the notation $\underline{g} := f_0$ and $\overline{g} := f_1$, that $K(\underline{g}) \leq V \leq K(\overline{g})$. The problem is then to find a policy f^* in the parametrized family $\{f_q, 0 \leq q \leq 1\}$ that *steers* the cost (2.2) to the value V, i. e., $K(f^*) = V$. If the mapping $q \rightarrow K(f_q)$ is *continuous*, this can be achieved by selecting f^* to be f_{q^*}, with the bias q^* being determined as the solution to the equation $K(f_q) = V, \quad 0 \leq q \leq 1$. Although most of the ideas discussed in this paper apply *mutatis mutandis* to this more general situation, the discussion will be carried out only in the context of constrained MDP's for sake of clarity.

3.3 A time-sharing implementation

Although the optimal mixing policy $\pi(p^*)$ of Section 2.4 has a very simple structure, it is *not* stationary and ergodic. Indeed, under such a policy $\pi(p^*)$, the *sample averages* corresponding to $K(\pi(p^*))$ do not satisfy the constraint since on a set of probability p^* (resp. $1 - p^*$), these limits will be $K(\overline{g})$ (resp. $K(\underline{g})$). This somewhat unappealing feature can be eliminated through the following *time-sharing* implementation of mixing policies. Assume the existence of a privileged state to which the system returns to infinitely often under each one of the policies $\underline{g}$ and $\overline{g}$, and define a cycle as the time T between consecutive visits to that state. Denote the expectation of a cycle duration under policies $\underline{g}$ and $\overline{g}$ by $E^{\underline{g}}(T)$ and $E^{\overline{g}}(T)$, respectively. For every p in the unit interval $[0,1]$, the mixing policy $\pi(p)$ has a *time-sharing* implementation $\alpha_{\mathrm{TS}}(p)$ which is now defined: Let $\tilde{p}$ be the element of $[0,1]$ uniquely defined through the relation

$$p = \frac{\tilde{p}E^{\overline{g}}(T)}{(1-\tilde{p})E^{\underline{g}}(T) + \tilde{p}E^{\overline{g}}(T)} \tag{3.2}$$

and consider two sequences of non-negative integers $\{\underline{n}_j\}_1^\infty$ and $\{\overline{n}_j\}_1^\infty$ with the property that

$$\lim_J n(J) = \infty \quad \text{and} \quad \lim_J \frac{\overline{n}(J)}{n(J)} = \tilde{p}, \tag{3.3}$$

where the notations

$$\underline{n}(J) := \sum_{j=1}^{J} \underline{n}_j, \quad \overline{n}(J) := \sum_{j=1}^{J} \overline{n}_j \quad \text{and} \quad n(J) := \underline{n}(J) + \overline{n}(J) \tag{3.4}$$

are used for every J in $I\!N$. The discrete-time axis is divided into contiguous *control frames*; the $(J+1)^{rst}$ such control frame starts upon completion of the $n(J)^{th}$ cycle and is made up of $\underline{n}_{J+1} + \overline{n}_{J+1}$ cycles. The policy $\alpha_{\mathrm{TS}}(p)$ is defined as the policy in $\mathcal{P}$ that during the J^{th} frame operates policy $\underline{g}$ for $\underline{n}_J$ cycles, and then policy $\overline{g}$ for $\overline{n}_J$ cycles, $J = 1, 2, \cdots$. Under the condition (3.3), well-known properties of first return times for Markov chains readily imply that

$$J(\alpha_{\mathrm{TS}}(p)) = \frac{(1-\tilde{p})E^{\underline{g}}(T)J(\underline{g}) + \tilde{p}E^{\overline{g}}(T)J(\overline{g})}{(1-\tilde{p})E^{\underline{g}}(T) + \tilde{p}E^{\overline{g}}(T)} = J(\pi(p)) \tag{3.5a}$$

and

$$K(\alpha_{\mathrm{TS}}(p)) = \frac{(1-\tilde{p})E^{\underline{g}}(T)K(\underline{g}) + \tilde{p}E^{\overline{g}}(T)K(\overline{g})}{(1-\tilde{p})E^{\underline{g}}(T) + \tilde{p}E^{\overline{g}}(T)} = K(\pi(p)) \tag{3.5b}$$

where the second equality is justified through (2.12) by the definition of $\tilde{p}$. In fact, the convergence (3.5) takes place for the sample averages as well. Note that if $\tilde{p}$ were *rational*, say of the form $\tilde{p} = \frac{\overline{n}}{\overline{n}+\underline{n}}$ for some integers $\underline{n}$ and $\overline{n}$, then the conditions (3.3) would be automatically satisfied upon choosing $\underline{n}_j = \underline{n}$ and $\overline{n}_j = \overline{n}$ for all j in $I\!N$.

The reader will readily see that the time-sharing implementation $\alpha_{\mathrm{TS}}(p^*)$ of the optimal mixing policy $\pi(p^*)$, satisfies the requirements (R1)-(R3) and thus solves the constrained problem (CP_V).

In Section 4, this approach is shown to be useful for identifying a solution to a multiple constraint problem.

3.4 A Certainty Equivalence implementation

A possible solution to the difficulties mentioned in Section 3.2 would be to estimate *directly* the optimal bias q^* and then use the *Certainty Equivalence Principle* at each step. More precisely, this suggests using a possibly *recursive* estimation scheme that generates a sequence of bias values $\{q(n)\}_1^\infty$ converging to q^*. At step n, the RV $q(n)$ constitutes an *estimate* of the bias value q^*, which is thus interpreted as the (conditional) probability of using $\overline{g}$ (given available information $H(n)$), and it is thus natural to select the control action $U(n)$ according to $f_{q(n)}$; the adaptive policy so generated by the sequence $\{q(n))\}_1^\infty$ is denoted by α.

There are as many such adaptive schemes as there are schemes for estimating the optimal bias value q^*. In each specific case, optimality of the adaptive policy α will be concluded if it can be established that policy α

(A1): yields (2.1) and (2.2) as limits,

(A2): meets the constraint, i. e., $K(\alpha) = V$, and

(A3): yields the same cost as f^*, i. e., $J(\alpha) = J(f^*)$.

This is done via a separate analysis and typically proceeds by showing that

$$\lim_n |f_{q(n)}(X(n)) - f^*(X(n))| = 0, \tag{3.6}$$

the convergence taking place under P^α either *almost surely* or *in probability*, a property which readily follows from the *(weak) consistency* of the estimation scheme. Under (3.6) and possibly additional structural model assumptions, a method of proof due to Mandl [19] can be extended to show that $J(\alpha) = J(f^*)$ and $K(\alpha) = K(f^*)$. Examples of this approach are given in Sections 4 and 5.

At this point, the reader may wonder as to how such an estimation scheme is selected.

(i): Sometimes, it is feasible to compute the optimal bias q^* as a function $q^*(\theta)$ of the external parameters θ. In that case, the designer may want to consider using the Certainty Equivalence Principle in conjunction with a parameter estimation scheme, say based on the *Maximum Likelihood* Principle. This approach is illustrated in Section 5 on a problem of flow control.

(ii): In many applications, the function $q \rightarrow K(f_q)$ turns out to be *continuous* and *strictly monotone*, say increasing for sake of definiteness. In that case, the search for q^* can be interpreted as finding the zero of the continuous, strictly monotone function $K(f_q) - V$ and this brings to mind ideas from the theory of *Stochastic Approximations* [24]. Here, this circle of ideas suggests generating a sequence of bias values $\{q(n)\}_1^\infty$ through the recursion

$$q(n+1) = [\ q(n) + a_n(V - c(X(n+1), U(n+1)))\ ]_0^1 \qquad n = 1, 2, \cdots \tag{3.7}$$

with $q(1)$ given in [0,1]. Here $[x]_0^1 := 0 \vee (x \wedge 1)$ for all x in $\mathbb{R}$ and the sequence of step sizes $\{a_n\}_1^\infty$ satisfies

$$0 < a_n \downarrow 0, \sum_{n=1}^{\infty} a_n = \infty, \sum_{n=1}^{\infty} |\ a_{n+1} - a_n\ | < \infty. \tag{3.8}$$

The corresponding policy α_{SA} is structurally simple and easy to implement on-line; this simplicity of implementation is derived from the fact that the difficult step of *directly* solving for q^* is completely bypassed.

4. OPTIMAL RESOURCE ALLOCATION

4.1 Model

Consider the following system of $K+1$ infinite-capacity queues that compete for the use of a single server: Time is slotted and the service requirement of each customer corresponds exactly to one time slot. At the beginning of each time slot, the controller gives priority to one of the queues. If the k^{th} queue is given service attention during that slot, then with probability μ_k the serviced customer (if any) completes service and leaves the system, while with probability $1-\mu_k$, the customer fails to complete service and remains in the queue. The arrival pattern is modelled as a *renewal* process, in that the batch sizes of customers arriving into the system in each slot are *independent* and *identically* distributed from slot to slot. Under these assumptions, the evolution of the system is fully described by an $I\!N^{K+1}$-valued process $\{X(n)\}_1^\infty$, with $X_k(n)$, $0 \leq k \leq K$, representing the number of customers in the k^{th} queue at the beginning of the slot $[n, n+1)$.

The mean number of customers arriving to the k^{th} queue is denoted by λ_k. Define the traffic intensity $\rho := \sum_{k=0}^{K} \frac{\lambda_k}{\mu_k}$ and assume henceforth that $\rho < 1$; this guarantees system stability [2].

For some mapping $c: I\!N^{K+1} \to I\!R$, the cost to be minimized is defined by

$$J(\pi) := \overline{\lim}_n \frac{1}{n} E^\pi \sum_{t=1}^{n} c(X(t)) \tag{4.1}$$

for every policy π in $\mathcal{P}$. A special case abundantly treated in the literature is the one where c is *linear* and *positive*, i.e. for all x in $I\!N^{K+1}$,

$$c(x) = \sum_{k=0}^{K} c_k x_k, \tag{4.2}$$

with $c_k \geq 0, 0 \leq k \leq K$. For this case, several authors have discussed the problem of selecting a service allocation strategy that minimizes (4.1) over the class $\mathcal{P}$ of all admissible service allocation strategies [3,5]. They all show the optimality of the μc-rule, i. e., the fixed priority assignment policy that orders the customer classes in increasing order of priority with the values $\mu_k c_k, 0 \leq k \leq K$.

4.2 Single constrained queue

In [21], Nain and Ross considered the situation where several types of traffic, e.g., voice, video and data, compete for the use of a single synchronous communication channel. They formulate this situation as a system of $K+1$ discrete-time queues that compete for the attention of a single server, and solve for the service allocation strategy that minimizes the long-run average of a linear expression in the queue sizes of the K customer classes $\{1, \cdots, K\}$ under the constraint that the long-run average queue size of the remaining customer class 0 does not exceed a certain value V. Thus for any policy π in $\mathcal{P}$, define $J(\pi)$ by (4.1) with c given by (4.2) where $c_0 = 0$, and pose

$$K(\pi) := \overline{\lim}_n \frac{1}{n} E^\pi \sum_{t=1}^{n} X_0(t). \tag{4.3}$$

Nain and Ross [21] extend some of the optimality results from [3,5] to show that if the constraint can be met in a non-trivial fashion, then an optimal policy with a very simple structure can be identified.

Theorem [21]: *If the problem is feasible, there exists a constrained optimal policy f^* in $\mathcal{F}$ which randomizes between two work-conserving static priority assignment policies.*

This result is derived through a Lagrangian argument in the following way: For $\gamma > 0$, let the mapping $b^\gamma : I\!N^{K+1} \to I\!R$ be given by $b^\gamma(x) = c(x) + \gamma x_0$ and pose

$$B^\gamma(\pi) := \overline{\lim}_n \frac{1}{n} E^\pi \sum_{t=1}^{n} b^\gamma(X(t)) \tag{4.4}$$

for every policy π in $\mathcal{P}$. The unconstrained MDP (LP_γ) is now defined as

$$(LP_\gamma): \quad \text{minimize } B^\gamma(\pi) \text{ over } \mathcal{P},$$

and solving the constrained problem (CP_V) reduces to finding a policy π^* in $\mathcal{P}$ which meets the requirements (R1)-(R3).

If the constraint is not satisfied while giving highest priority to the constrained queue (i. e., the 0^{th} queue), then there is no solution. On the other hand, if the constraint is satisfied while giving lowest priority to the constrained queue and ordering the other queues according to the μc-rule, then this policy is optimal for the constrained problem (CP_V). The proof in the remaining case is available in [21], the main idea being that each one of the problems (LP_γ) is solved by a fixed priority assignment policy which admits a description as the μc-rule based on $\mu_0\gamma$ and $\mu_k c_k$, $1 \leq k \leq K$.

The dynamics of this problem were generalized by Nain and Ross to a semi-Markov decision process [22]. Another generalization, given by Altman and Shwartz [1], involves a more general constraint (4.3) associated with an instantaneous cost $d\colon I\!N^{K+1} \rightarrow I\!R$ which is also an arbitrary linear function with positive coefficients (as in (4.2)).

Theorem [1]: *If the problem is feasible, then there is a constrained optimal policy f_{q^*} which randomizes between two work-conserving static priority assignment policies.*

4.3 Single constraint - An adaptive implementation

Even with (4.3), the function $q \rightarrow K(f_q)$ is not easy to compute, in spite of the linearity of the instantaneous cost. Indeed, as pointed out by Nain and Ross [21], computing the quantity $K(f_q)$ amounts to studying a coupled-processor problem whose solution can be obtained via a reduction to a *Riemann-Hilbert* problem [6].

The stochastic approximation algorithm that generates the sequence of bias estimates $\{q(n)\}_1^\infty$ here takes the special form

$$q(n+1) = \left[\ q(n) + a_n(V - X_0(n+1))\ \right]_0^1 \qquad n = 1, 2, \cdots \quad (4.5)$$

with $q(1)$ given in [0,1].

For $K = 1$, there are only two fixed priority assignment policies, and therefore *no a priori* knowledge of the various statistics is required in order to implement this algorithm. As such, the proposed policy is implementable and constitutes an adaptive policy in the restricted technical sense understood in the literature on the non-Bayesian adaptive control problem for Markov chains [11]. However, for $K > 1$, α_{SA} is implementable only if $\underline{g}$ and $\overline{g}$ have been determined.

The basic results assume finite third moments on the initial queue sizes and on the statistics of the arrival pattern. Under these additional conditions,

Theorem [30]: *The sequence of biases $\{q(n)\}_1^\infty$ converges in probability (under $P^{\alpha_{\rm SA}}$) to the optimal bias q^*.*

The derivation of this result makes combined use of results by Kushner and Shwartz [14] on the weak convergence of Stochastic Approximations via ODE methods and on moment estimates obtained for the queue size process in [30]. The condition (3.6) now holds and implies

Theorem [30]: *The policy $\alpha_{\rm SA}$ solves the constrained optimization problem (CP_V) with $J(\alpha_{\rm SA}) = J(f^*)$ and $K(\alpha_{\rm SA}) = K(f^*)$.*

4.4 Multiple constraint - A time-sharing implementation

In this section, a special version of the multiple constraint problem (CP_V) is studied. For every policy π in $\mathcal{P}$, pose

$$K^i(\pi) := \overline{\lim}_n \frac{1}{n} E^\pi \sum_{t=1}^n X_i(t), \quad 0 \le i \le K, \tag{4.6}$$

and for every vector $V = (V_1, \cdots, V_K)$ in $I\!R^K$, consider the constrained problem

$$(CP_V): \quad \text{minimize } K^0(\pi) \text{ over } \mathcal{P}_V, \tag{4.7}$$

with $\mathcal{P}_V$ defined by (3.1).

There are exactly $L := (K+1)!$ non-idling policies in $\mathcal{G}$ which act as fixed priority assignments, the l^{th} such policy being denoted throughout by g_l, $1 \le l \le L$. An element $p = (p_1, \cdots, p_L)$ in $I\!R^L$ lies in the L-dimensional simplex whenever $0 \le p_l \le 1$ for all $1 \le l \le L$, with $\sum_{l=1}^L p_l = 1$. The *mixing* policy $\pi(p)$ associated with any element p in the L-dimensional simplex is defined through a procedure that generalizes the one discussed in Section 2.4: Consider an L-sided coin that yields the l^{th} side with probability p_l, $1 \le l \le L$ and throw the coin exactly once at the beginning of times, before starting to operate the system. The policy $\pi(p)$ is the one that operates according to the fixed priority policy g_l if the outcome of the throw is the l^{th} side, $1 \le l \le L$. For every p in the L-dimensional simplex, the relations

$$K^i(\pi(p)) = \sum_{l=1}^L p_l K^i(g_l), \quad 0 \le i \le K, \tag{4.8}$$

hold true and suggest the following *Linear Program* (LP), where

$$\text{Minimize} \quad \sum_{l=1}^{L} z_l K^0(g_l) \tag{4.9a}$$

subject to the constraints

$$\sum_{l=1}^{L} z_l K^i(g_l) \leq V_i, \quad 0 \leq i \leq K, \tag{4.9b}$$

and

$$0 \leq z_l \leq 1, 1 \leq l \leq L, \quad \text{and} \quad \sum_{l=1}^{L} z_l = 1. \tag{4.9c}$$

The relationship between the Linear Program (LP) and the original constrained problem (CP_V) is formalized in the following result.

Theorem [1]: *If the Linear Program (LP), has a solution, say p^*, then the policy $\pi(p^*)$ solves the multiple constraint problem (CP_V). Conversely, if problem (CP_V) has a solution, then it can always be implemented by a mixing policy.*

The solution of problem (CP_V) via mixing policies has the clear advantage of requiring only the solution of the Linear Program (LP); this is to be contrasted with the difficult queueing problem that is required to find the optimal randomized policy [21]. A dynamic or time-sharing implementation can also be provided in the spirit of Section 3.3. Here the empty state is taken to be the privileged state and a cycle thus coincides with a *busy* cycle for the queueing system. For sake of *brevity*, the discussion is restricted to the case where the element p in the L-dimensional simplex has all its components *rational* with $p_l = \frac{n_l}{|n|}$, $1 \leq l \leq L$, for some $n = (n_1, \cdots, n_L)$ in $I\!N^L$, where $|n| = \sum_1^L n_l$. Define the policy $\tilde{\pi}(n)$ in $\mathcal{P}$ as the one that operates the fixed priority assignment g_l over n_l cycles. With the help of results on busy periods [2,29], it is a simple exercise to show, in analogy with (3.5), that the relations

$$K^i(\alpha_{\text{TS}}(p)) = \sum_{l=1}^{L} \frac{n_l}{|n|} K^i(g_l), \quad 0 \leq i \leq K, \tag{4.10}$$

hold true. The more general situation can be treated exactly as in Section 3.3.

5. OPTIMAL FLOW CONTROL

5.1 Model and constrained problems

Consider the following flow control model for discrete-time $M|M|1$ queueing systems [17,18]: At the beginning of each time slot, the controller decides either to admit or reject the arrivals during that slot. An admitted customer joins the queue while a rejected customer is immediately lost. A customer (if any) that completes service in a slot leaves the system at the end of that slot with probability μ, and fails to complete service in that slot with probability $1-\mu$, in which case it remains at the head of the line to await service in the next slot. The arrival pattern is modelled as a *Bernoulli* sequence with parameter λ, *independent* of the service process which is modelled as another *Bernoulli* sequence with parameter μ. Under these assumptions, the evolution of the system is fully described by an $I\!N$-valued process $\{X(n)\}_1^\infty$, with $X(n)$ representing the number of customers in the queue at the beginning of the slot $[n, n+1)$.

The problem considered here is formulated as the search for a policy that maximizes the throughput under the constraint that the long-run average queue size does not exceed a certain value V, where the throughput and the average queue size are expressed as

$$J(\pi) := \underline{\lim}_n \frac{1}{n} E^\pi \sum_{t=1}^{n} \mu 1(X(t) \neq 0) \tag{5.1}$$

and

$$K(\pi) := \overline{\lim}_n \frac{1}{n} E^\pi \sum_{t=1}^{n} X(t), \tag{5.2}$$

respectively, for every admissible policy π on $\mathcal{P}$.

A *threshold* policy is a Markov stationary policy in $\mathcal{F}$, with a simple structure determined by two parameters L and q in $I\!N$ and $[0,1]$, respectively, whence a threshold policy is denoted hereafter by (L, q). According to the threshold policy (L, q), an incoming customer is admitted or rejected wether the queue size is $< L$ or $> L$; if the queue size is exactly L, a biased coin with bias q is flipped and the outcome then determines whether or not the incoming customer can access to the queue. The adopted convention interprets q as the (conditional) probability of accepting an incoming customer.

The policy that admits every single customer is denoted by $(\infty, 1)$. If $K((\infty,1)) \leq V$, then the problem (CP_V) corresponding to (5.1)-(5.2) is trivially solved by $(\infty, 1)$. On the other hand, if $K((\infty,1)) > V$, then the problem has a non-trivial solution, which can be shown to be of threshold type. This result is derived again through a Lagrangian argument: For $\gamma > 0$, let the mapping $b^\gamma : I\!N \rightarrow I\!R$ be given by $b^\gamma(x) = \mu 1(x \neq 0) - \gamma x$ and define for every policy π in $\mathcal{P}$, $B^\gamma(\pi)$ as in (2.5). The unconstrained MDP (LP_γ) is now defined as Section 2, and as in previous instances where this technique was used, solving the constrained problem (CP_V) reduces to finding a policy π^* in $\mathcal{P}$ which meets the requirements (R1)-(R3).

The unconstrained Lagrangian problem (LP_γ) can be solved through a tedious Dynamic Programming argument that shows *concavity* of the corresponding value function.

Theorem [17]: *For every $\gamma > 0$, there exists a threshold policy (L_γ, q_γ) that solves the unconstrained Lagrangian problem (LP_γ). Moreover, for each threshold value L in $I\!N$, there always exists $\gamma(L) > 0$ so that any threshold policy (L,q), with q arbitrary in* $[0,1]$, *solves the Lagrangian problem $(LP_{\gamma(L)})$.*

Threshold policies are thus unconstrained optimal. For any threshold policy (L,q), the quantities (5.1) and (5.2) exist as *limits.* Moreover, $K((L,q))$ *increases* as L increases, whereas for fixed L in $I\!N$, the mapping $q \rightarrow K((L,q))$ is *continuous* and *strictly monotone increasing* on [0,1]. Since $K((\infty,1)) > V$, there exists L^* in $I\!N$ such that $K((L^*,0)) < \mathrm{V} \leq K((L^*,1))$, and the continuity and the strict monotonicity of $q \rightarrow K((L^*,q))$ then imply the existence of q^* such that $K((L^*,q^*)) = V$.

Theorem [17]: *If the constrained problem (CP_V) has a non-trivial solution, it can be taken to be a threshold policy $f^* = (L^*, q^*)$.*

The quantities L^* and q^* are determined by the arrival and service rates λ and μ, and by the constraint value V; they can be computed as the unique solution to the equation

$$K((L,q)) = V, \quad L = 0, 1, \cdots \text{and} \quad 0 \leq q \leq 1. \tag{5.3}$$

This result represents the discrete-time analog of results obtained by Lazar [15]. In contrast with the competing queue problem, a *closed-form* expression is available here for the quantity $K((L,q))$ for all L in $I\!N$ and q in [0,1]. However, as in Section 4, it is still reasonable to seek on-line implementations of such a threshold policy.

5.2 A Stochastic Approximation implementation

As in Section 3.4, the two policies $\underline{g} = (L^*, 0)$ and $\overline{g} = (L^*, 1)$ are assumed available, which presumes only knowledge of L^*. The following stochastic approximation algorithm generates the sequence of bias estimates $\{q(n)\}_1^\infty$ through the recursion

$$q(n+1) = [\ q(n) + a_n(V - X(n+1))\ ]_0^1 \qquad n = 1, 2, \cdots (5.4)$$

with $q(1)$ given in [0,1], where in addition to the conditions (3.8), the step sizes $\{a_n\}_1^\infty$ satisfy $\sum_{i=1}^\infty a_i^2 < \infty$. The RV $q(n)$ constitutes an estimate of the bias value q^* and is interpreted as the conditional probability of using $\overline{g}$, or equivalently, of giving admission to a potential customer during the slot $[n, n+1)$ when the queue size $X(n)$ is equal to L^*. With this scheme, the control action to be implemented is simply generated according to the optimal threshold policy f^* when the queue size is not equal to L^*.

The key result is obtained under a fourth moment assumption on the initial queue size, and is proved using ideas proposed by Metivier and Priouret [20] on the almost sure convergence of Stochastic Approximation algorithms.

Theorem [18]: *The sequence of biases $\{q(n)\}_1^\infty$ converges almost surely (under $P^{\alpha_{\mathrm{SA}}}$) to the optimal bias q^*.*

The condition (3.6), which is now seen to hold, can then be used to show that

Theorem [18]: *The policy α_{SA} solves the constrained optimization problem (CP_V) with $J(\alpha_{\mathrm{SA}}) = J(f^*)$ and $K(\alpha_{\mathrm{SA}}) = K(f^*)$.*

5.3 The time-sharing implementation

Since the quantity $K((L, q))$ is computable for all values of its arguments, the values L^*, $K(\underline{g})$ and $K(\overline{g})$ are thus readily available. The optimal mixing parameter p^* can then be immediately evaluated, and the optimal mixing policy $\pi(p^*)$ considered in Section 2.4 is thus easily implementable. Here, the threshold nature of the policies $\underline{g}$ and $\overline{g}$ suggests level L^* as a privileged state, and a cycle is thus defined as the time duration between consecutive visits to level L^*. The time-sharing implementation $\alpha_{\mathrm{TS}}(p^*)$ corresponding to $\pi(p^*)$ is defined as in Section 3.3 and provides an easy way to implement an optimal constrained policy. The discussion is similar to the one of Section 3.3 and will be omitted.

5.4 An indirect adaptive control implementation

All previous implementation schemes require the knowledge of the optimal threshold L^*. In case the external parameters, λ and μ, are unknown, the parameter L^* is certainly not available and all previously considered policies are thus not implementable in their given form.

In such cases, it is natural to consider a scheme that uses Certainty Equivalence principle in conjunction with maximum likelihood estimators. The sequence of estimates $\{(\lambda(n), \mu(n))\}_1^\infty$ of the true parameter (λ, μ) is generated based on all the past available information by invoking the principle of maximum likelihood, and the sequence $\{(L(n), q(n))\}_1^\infty$ is then determined by the estimates $\lambda(n)$ and $\mu(n)$ by solving the equation (5.3). In case there is no solution to the equation (5.3) for a pair $(\lambda(n), \mu(n))$, simply set $L(n) = \infty$ and $q(n) = 1$. The control action to be implemented in the slot $[n, n+1)$ is then generated according to $(L(n), q(n))$, and the corresponding adaptive policy is denoted by α_{ML}.

The estimation procedure relies on a information pattern $I(n)$, richer than $H(n)$ and given by

$$I(n) = \{X(1), U(t), A(t), B(t), 1 \le t < n\} \qquad n = 1, 2, \cdots \quad (5.5)$$

where the $\{0,1\}$-valued RV's $U(n), A(n)$ and $B(n)$ represent the control action implemented in the slot $[n, n+1)$, the arrival during that slot and service completion at the end of that slot, respectively.

Using the Strong Law of Large Numbers and results from the theory of Large Deviations, the appropriate version of condition (3.6) is shown to again take place [18], the rate of convergence being *exponentially* fast; this last fact is crucial to the basic result, which is obtained under a fourth moment assumption on the initial queue size.

Theorem [18]: *The policy α_{ML} solves the constrained optimization problem (CP_V) with $J(\alpha_{\mathrm{ML}}) = J(f^*)$ and $K(\alpha_{\mathrm{ML}}) = K(f^*)$.*

It is not clear at this time on how to design *direct* adaptive control schemes when the optimal threshold L^* is not available. A stochastic approximation algorithm for generating recursively a sequence of estimates for L^* similar to the one used for q^* naturally comes to mind; however, the corresponding scheme fails to work owing to its sensitivity to variations of the *integer*-valued threshold [18].

REFERENCES

[1] E. Altman and A. Shwartz, "Optimal priority assignment with general constraints", submitted, 24th Allerton Conference on Communication, Control and Computing, October 1986.

[2] J. S. Baras, A. J. Dorsey and A. M. Makowski, "Discrete time competing queues with geometric service requirements: stability, parameter estimation and adaptive control", under revision, SIAM J. Control Opt.

[3] J. S. Baras, D.-J. Ma and A. M. Makowski, "K competing queues with geometric service requirements and linear costs: The μc-rule is always optimal", Systems & Control Letters, vol. 6, pp. 173-180 (1985).

[4] Beutler and Ross, "Time-average optimal constrained semi-Markov decision processes", Adv. Appl. Prob. vol. 18, pp. 341-359 (1986).

[5] C. Buyyukkoc, P. Varaiya and J. Walrand, "The $c\mu$-rule revisited", Adv. Appl. Prob. vol. 17, pp. 234-235 (1985).

[6] G. Fayolle and P. Iasnogorodski, "Two coupled processors: The reduction to a Riemann-Hilbert problem", Z. Wahr. verw. Gebiete, vol. 47, pp. 325-351 (1979).

[7] G. C. Goodwin and K. W. Sin, *Adaptive Filtering, Prediction and Control*, Prentice-Hall, Englewood Cliffs, NJ., 1984.

[8] B. Hajek, "Optimal control of two interacting service stations", IEEE Trans. Auto. Control, vol. AC-30, pp. 68-76 (1985).

[9] D. Heyman and M. Sobel, *Stochastic Models in Operations Research, Volume II: Stochastic Optimization*, MacGraw-Hill, New York, 1984.

[10] L. Kleinrock, *Queueing Systems, Volume II: Computer Applications* , John Wiley & Sons, New York, 1976.

[11] P. R. Kumar, "A survey of some results in stochastic adaptive control", SIAM J. Control Opt. vol. 23, pp. 329-380 (1985).

[12] P. R. Kumar and W. Lin, "Optimal control of a queueing system with two heterogeneous servers", IEEE Trans. Auto. Control, vol. AC-29, pp. 696-703 (1984).

[13] H. J. Kushner, *Introduction to Stochastic Control*, Holt, Rinehart and Winston, New York, 1971.

[14] H. J. Kushner and A. Shwartz, "An invariant measure approach to the convergence of Stochastic Approximations with state-dependent noise", SIAM J. Control Opt. Vol. 22, pp.13-27 (1984).

[15] A. A. Lazar, "The throughput time delay function of an $M|M|1$ queue", IEEE Trans. Info. Theory, vol. 29, pp. 914-918 (1983).

[16] S. Lippman, "Applying a new device in the optimization of exponential queueing systems", Oper. Res. vol. 23, pp. 687-710 (1975).

[17] D.-J. Ma and A. M. Makowski, "A simple problem of folw control I: Optimality results", in preparation (1986).

[18] D.-J. Ma and A. M. Makowski, "A simple problem of folw control II: Implementation of threshold policies", in preparation (1986).

[19] P. Mandl, "Estimation and control in Markov chains", Adv. Appl. Prob. vol. 6, pp. 40-60 (1974).

[20] M. Metivier and P. Priouret, "Applications of a Kushner and Clark lemma to general classes of stochastic algorithms", IEEE Trans. Info. Theory, vol. 30, pp. 140-150 (1984).

[21] P. Nain and K. W. Ross, "Optimal priority assignment with hard constraint", Rapports de Recherche No. 459, INRIA - Rocquencourt, France (November 1985).

[22] P. Nain and K. W. Ross, "Optimal multiplexing of heterogeneous traffic with hard constraint", Proceedings of the Joint Performance '86 and ACM Sigmetrics Conference, Raleigh, North Carolina, May 1986.

[23] M. Reiser, "Performance evaluation of data communication systems", Proc. IEEE, vol. 70, no 12, pp.171-196 (1982).

[24] H. Robbins and S. Monro, "A stochastic approximation method", Ann. Math. Stat. vol. 22, pp. 400-407 (1951).

[25] Z. Rosberg, P. V. Varaiya and J. Walrand, "Optimal control of service in tandem queues", IEEE Trans. Auto. Control, vol. AC-27, pp. 600-610 (1982).

[26] K. W. Ross, *Constrained Markov Decision Processes with Queueing Applications*, Ph.D. thesis, Computer, Information and Control Engineering, University of Michigan, 1985.

[27] D. Serfozo, "An equivalence between discrete and continuous time Markov decision processes", Oper. Res. vol. 23, pp. 616-620 (1979).

[28] A. Shwartz and A. M. Makowski, "An optimal adaptive scheme for two competing queues with constraints", 7th International Conference on Analysis and Optimization of Systems, pp. 515-532, Antibes, France, 1986.

[29] A. Shwartz and A. M. Makowski, "Adaptive policies for a system of competing queues I: Convergence results for the long-run average cost", in preparation (1986).

[30] A. Shwartz and A. M. Makowski, "Adaptive policies for a system of competing queues II: Implementable schemes for optimal server allocation", in preparation (1986).

[31] S. Stidham, Jr., "Optimal control of admission to a queueing system", IEEE Trans. Auto. Control, vol. AC-30, pp. 705-713 (1985).